

Claims

What is claimed is:

1. A method for converting a document from a first format to a second format, the method comprising:

5 dividing a document having a first structure into blocks;
 creating a list having entries, individual entries of the list containing content of associated ones of the blocks;
 providing a database including a data structure associated with the document, the data structure specifying a manner of displaying at least one of the entries;
10 inserting entries of the list into the data structure to form an output file.

2. The method for converting a document from a first format to a second format according to claim 1, further comprising transmitting at least a portion of the output file over a network to a client device.

15

3. The method for converting a document from a first format to a second format according to claim 1 wherein the creating a list further comprises assigning a classification to individual list entries.

20 4. The method for converting a document from a first format to a second format according to claim 3, further comprising merging list entries having the same classification.

5. The method for converting a document from a first format to a second format according to claim 1 wherein the creating a list further comprises re-ordering the list according to the content of individual list entries.
- 5 6. The method for converting a document from a first format to a second format according to claim 1 wherein the output file contains sub-documents and a table of contents page listing the labels, wherein individual sub-documents are associated with individual labels.
- 10 7. The method for converting a document from a first format to a second format according to claim 1, further comprising:
 - extracting a structure of the document to form an extracted data structure associated with the document;
 - modifying the extracted data structure;

15 storing the modified extracted data structure as the data structure.
8. The method for converting a document from a first format to a second format according to claim 7, wherein the modifying the extracted data structure further comprises adding labels to the extracted data structure.
- 20 9. The method for converting a document from a first format to a second format according to claim 8, wherein the modifying the extracted data structure further comprises removing a portion of the extracted data structure.

10. The method for converting a document from a first format to a second format according to claim 8, wherein the modifying the extracted data structure further comprises removing a portion of the extracted data structure from a first location within the extracted data structure and adding the portion of the extracted data structure at a 5 second location within the extracted data structure.

11. The method of converting a document from a first format to a second format according to claim 7, wherein the document comprises an HTML document, an XML document, or a PDF document.

10

12. A method for converting a document from a first format to a second format, the method comprising:

extracting a structure of a first document to form a first data structure;
modifying the first data structure to form a second data structure;
15 extracting content of a second document;
inserting the content of the second document into the second data structure, the content of the second document being different from content of the first document.

13. The method for converting a document from a first format to a second format 20 according to claim 12, wherein the modifying the first data structure further comprises deleting a portion of the first data structure.

14. The method for converting a document from a first format to a second format according to claim 12, wherein the modifying the first data structure further comprises adding a label to a portion of the first data structure.

5 15. The method for converting a document from a first format to a second format according to claim 12, wherein the first and second documents are web pages.

16. A method for converting a document from a first format to a second format, the method comprising:

10 extracting a first data structure from a first document, content of the first document being stored in nodes of the data structure;
assigning a label to the nodes of the first data structure that store the content of the document based on the content stored in the nodes;
generating a one-dimensional list of the nodes that include the content of the
15 document.

17. The method for converting a document from a first format to a second format according to claim 16,

20 providing a database including a second data structure associated with the first document, the second data structure specifying a manner of displaying at least one of the entries;
inserting entries of the list into the data structure to form an output file.

18. The method for converting a document from a first format to a second format

according to claim 17, wherein the providing a database further comprises:

extracting a structure of a second document to form a second data structure, the

second document having a same structure as the first document;

5 modifying the second data structure to form a third data structure;

storing the third data structure in a database.

19. The method for converting a document from a first format to a second format

according to claim 18, wherein the modifying further comprises removing at least one

10 portion of the second data structure.

20. The method for converting a document from a first format to a second format

according to claim 18, wherein the modifying further comprises moving at least one

portion of the second data structure from a first location within the second data structure

15 to a second location within the data structure.

21. A computer readable medium comprising program instructions for:

dividing a document having a first structure into blocks;

creating a list having entries, individual entries of the list containing content of

20 associated ones of the blocks;

providing a database including a data structure associated with the document, the

data structure specifying a manner of displaying at least one of the entries;

inserting entries of the list into the data structure to form an output file.

22. A computer readable medium comprising program instructions for:
 - extracting a structure of a first document to form a first data structure;
 - modifying the first data structure to form a second data structure;
 - extracting content of a second document;
- 5 inserting the content of the second document into the second data structure, the content of the second document being different from content of the first document.
23. A computer readable medium comprising program instructions for:
 - extracting a first data structure from a first document, content of the first
 - 10 document being stored in nodes of the data structure;
 - assigning a label to the nodes of the first data structure that store the content of the document based on the content stored in the nodes;
 - generating a one-dimensional list of the nodes that include the content of the document.